DOCUMENT RESUME

ED 295 455 FL 016 629

AUTHOR SantaVicca, Edmund F.
TITLE Languages and the Information Professions: Implications for the Storage, Management, and Retrieval of Online Information.
PUB DATE Apr 86
NOTE 20p.; In: Proceedings of the Fifth Annual Eastern Michigan University Conference on Languages for Business and the Professions (see FL 016 586).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Access to Information; Databases; Educational Needs; Efficiency; Indexing; *Information Retrieval; *Information Scientists; Information Sources; *Information Storage; *Languages for Special Purposes; Librarians; *Online Searching; Second Language Instruction; Second Languages; Thesauri; Vocabulary

ABSTRACT

As more professionals come to rely on computer-based online information systems, knowledge of foreign languages and cultures plays a crucial role in the accurate storage, management, and retrieval of needed information. In the storage and management of information, foreign language knowledge is important in word choice and thesaurus control, indexing, cross-referencing, and arranging points of access. Online searching of databases for materials written, recorded, or described in a foreign language requires simultaneous use of three languages: the native language, the foreign language in question, and the language of the online system. Deficiency in any of these languages will result in less efficient or less effective information retrieval. Knowledge of foreign languages is increasingly important as information science advances. Without this knowledge, the information-seeker will be severely limited. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED295455

LANGUAGES AND THE INFORMATION PROFESSIONS: IMPLICATIONS FOR THE STORAGE, MANAGEMENT AND RETRIEVAL OF ONLINE INFORMATION

Dr. Edmund F. SantaVicca
Humanities Reference Bibliographer

University Libraries
Cleveland State University
1860 East 22nd Street
Cleveland, Ohio 44115 U.S.A.

(216) 687-2365

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

C. J. Voight

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FL016629

Edmund F. SantaVicca

Cleveland State University

Languages and the Information Professions: Implications for the Storage, Management and Retrieval of On-Line Information

Computer hardware and software systems continue to dominate the officescape of the 1980s. Their presence and role in the international access and sharing of information, through telecommunications and other on-line systems, is likely to increase significantly into the next century. This will hold true in at least three major information-centered sectors: Academe; the business, finance and corporate sector; other professional and paraprofessional sectors. As more business professionals, lawyers, doctors, educators, librarians and information scientists come to rely upon computer-based on-line information systems, knowledge of foreign languages and cultures plays a crucial role in the accurate storage, management and retrieval of information needed to meet the individual professional's information needs. Through examination of three types of on-line systems, and the parameters of data access within each system, we can better understand the importance of foreign languages in the present and future on-line setting.

At present, there are three general types of on-line information systems which the professional may use in searching

for computer-maintained information. The first type is comprised of commercial data bases, available through contractual agreement, that provide the professional with a variety of specialized data bases, a variety of data formats, and a multitude of access points to those formats. These bases number well into the hundreds, and may be accessed individually or simultaneously, depending upon the information needs of the user. Most of these on-line data bases have corresponding print formats, although their inherent value is that they can provide access to information more quickly than their corresponding print formats; and they can provide information which is more up-to-date than the corresponding print format. Examples of systems of commercial data bases include Lockheed's DIALOG, System Development Corporation's ORBIT, Bibliographic Retrieval Service (BRS), H.W. Wilson's WILSONLINE, MEDLINE, etc.

A second type of on-line system is represented by contractually-shared catalogs. In most instances, these are shared catalogs of libraries, museums, special collections, information centers and corporate data banks throughout the country. These systems can be local, regional, state-wide, national or international; and they usually include a variety of cataloged formats, with the exception of periodical articles. In comparison with most commercial data bases, these shared catalogs generally do not provide as many access points to needed information; and they are limited to data supplied by participating libraries or other data-generating entities. The prime example of this type of shared catalog is OCLC (Online Computer Library Center).

A third type of on-line system is represented by various software and hardware configurations designed for the in-house microcomputer environment. Such systems may or may not be designed to tie in, or interface, with other on-line systems, including local area networks (LAN), word processors, etc. Although such systems abound with great variety, the example which will be used here for the sake of illustration is the dBASE III software package, developed by Ashton-Tate.

To begin, let us examine sample entries from each of these three different systems: DIALOG (as an example of a commercial data base), OCLC, and dBASE III. In the over two hundred data bases available through DIALOG, some data records are bibliographic while other are non-bibliographic. Examples of both kinds of record are presented in Figures 1 and 2, respectively. Figure 1 presents an on-line bibliographic data record as it has been entered into the data base: ERIC, File 1, updated monthly. Figure 2 presents an on-line non-bibliographic data record as it has been entered into another of the bases available through DIALOG: TRINET COMPANY DATABASE, File 532, reloaded quarterly.

Notice that the information conveyed in Figure 1 includes an accession/document number; original title; translated title; author; publisher; date; pagination; language of text; availability in hard copy; language of abstract; document type; indication of original announcement; English-language abstract; and descriptors and identifiers assigned to the original work.

Figure 1

Commercial data base: bibliographic record: ERIC, File 1, updated monthly

ED122600 FL007414

LES CONNAISSANCES LINGUISTIQUES DES CADRES: LE POINT DE VUE DE LA DIRECTION DES ENTREPRISES. LES BESOINS LANGAGIERS DES FUTURS CADRES D'ENTREPRISE. (LANGUAGE SKILLS OF TRAINED PERSONNEL: THE VIEWPOINT OF BUSINESS MANAGEMENT. THE LANGUAGE NEEDS OF FUTURE BUSINESS MANAGERS).

HATRY, PAUL

UNIVERSITE LIBRE DE BRUXELLES (BELGIUM). INSTITUT DE PHONETIQUE.

NOV 1974 12P.; IN FRENCH

EDRS PRICE - MF01/PC01 PLUS POSTAGE.

LANGUAGE: ENGLISH

DOCUMENT TYPE: RESEARCH REPORT (143)

JOURNAL ANNOUNCEMENT: RIESEP76

THE FIRST PART OF THIS PAPER OUTLINES THE FACTORS THAT APPEAR TO INFLUENCE THE LANGUAGE ATTITUDES OR LANGUAGE BEHAVIOR OF BUSINESS DIRECTORS AND MANAGERIAL STAFF IN BELGIUM. AGE AND THE ECONOMIC AND POLITICAL SURROUNDINGS APPEAR TO BE THE PRINCIPAL FACTORS. THE YOUNGER INDIVIDUALS ARE, THE MORE POSITIVE ARE THEIR ATTITUDES ABOUT THE LANGUAGE DIVERSITY THEY ENCOUNTER AND THE NEED FOR LANGUAGE SKILLS IN BUSINESS. THE INTERNATIONALIZATION OF THE BELGIAN ECONOMY AND THE MULTINATIONALIZATION OF BUSINESS INFLUENCE DIRECTORIAL STAFF TO REQUIRE LANGUAGE SKILLS OF THEIR EMPLOYEES, PARTICULARLY WHEN THE NATIONAL LANGUAGE, SUCH AS IN HOLLAND AND THE SCANDINAVIAN COUNTRIES, IS NOT AN INTERNATIONAL LANGUAGE. THE CREATION OF INTERNATIONAL INSTITUTIONS CREATES FURTHER INCENTIVE. INDUSTRIES REQUIRING LANGUAGE SKILLS, SUCH AS THE AUTO INDUSTRY AND THE INDUSTRIES OF CHEMISTRY, OIL, AND STEEL, BANKING AND TOURISM, ARE CONTRASTED TO THOSE IN WHICH ONE'S MOTHER TONGUE IS SUFFICIENT, SUCH AS AGRICULTURE, THE FOOD INDUSTRY, UTILITIES AND CONSTRUCTION. THERE IS ALSO A NEED FOR THE ADMINISTRATIVE STAFF IN THE PUBLIC SECTOR TO KNOW MORE THAN ONE LANGUAGE IN ORDER TO FIT EFFECTIVELY INTO THE CONTEXT OF INTERNATIONAL BUSINESS. THE SECOND PART OF THE PAPER DESCRIBES THE HISTORY AND CHANGING PRIORITIES OF A BELGIAN BUSINESS SCHOOL, L'ECOLE DE COMMERCE SOLVAY. (CLK)

DESCRIPTORS: ADMINISTRATION; *ADMINISTRATORS; *BUSINESS COMMUNICATION; ECONOMIC FACTORS; *JOB SKILLS; *LANGUAGE ATTITUDES; LANGUAGE PROFICIENCY; *LANGUAGE SKILLS; *PERSONNEL NEEDS; POLITICAL INFLUENCES; PROGRAM DESCRIPTIONS; SECOND LANGUAGE LEARNING

IDENTIFIERS: *BELGIUM

Figure 2

Commercial data base: non-bibliographic record: TRINET

COMPANY DATABASE, File 532, reloaded quarterly

0222930 TRINET NUMBER: 99006233
TIFFANY INDUSTRIES INC
1055 CORP SQ DR
ST LOUIS, MO 63132

TELEPHONE: 314-991-1700
COUNTY: ST LOUIS IND CY
MSA: 7040 (ST. LOUIS, MISSOURI-ILLINOIS)
TICKER SYMBOL: TFY

NUMBER OF EMPLOYEES: 467
NUMBER OF ESTABLISHMENTS: 5

MANUFACTURING SALES ($): 28,900,000
NONMANUFACTURING SALES ($): 0
NON-US SALES ($): 900,000
TOTAL SALES ($): 29,800,000

PRIMARY 4 DIGIT SIC:
3448 (PREFABRICATED METAL BUILDINGS) SALES ($): 12,600,000

PRIMARY 3 DIGIT SIC:
344 (FABRICATED STRUCTURAL METAL PROD) SALES ($): 12,600,000

PRIMARY 2 DIGIT SIC:
34 (FABRICATED METAL PRODUCTS) SALES ($): 12,600,000

SECONDARY SIC(S):
2522 (METAL OFFICE FURNITURE) SALES ($): 10,100,000
3523 (FARM MACHINERY & EQUIPMENT) SALES ($): 6,200,000

THIS IS:
A PUBLIC COMPANY

One does not need, however, all of the information included in order to retrieve the record from the data base. Access to the record may be through descriptors or identifiers, other subject terms which may appear in the abstract, accession number, author, publisher, journal title (where applicable), publication year, or other access points which may be idiosyncratic to the individual file. In the case of ERIC, the data record requested can be printed in any of eight different formats: accession number only; full record except abstract; bibliographic citation; a file-dependent format; full record including abstract (Figure 1); title only; bibliographic citation and abstract; or title and indexing. Due to recent enhancements with the DIALOG system, and through skillful use of Boolean logic, access points to the data record can be combined, broadened or narrowed to assure optimum retrieval of desired information.

In comparison, the on-line non-bibliographic data record presented in Figure 2, from the TRINET COMPANY DATABASE, includes accession number, TRINET number, company name, company address, company telephone, county name, Metropolitan Statistical Area code and name, Ticker symbol, number of employees, number of establishments; manufacturing, non-manufacturing, non-United States, and total sales; primary four-digit, three-digit and two-digit Standard Industrial Classification codes; secondary SIC codes; and indication as to whether the company is public or private. Similar to ERIC, access to the data record can be

had through any of the various fields represented in the record; and, similar to ERIC, the user may request any one of a variety of formats: DIALOG accession number; company name, address, telephone number, SIC codes and descriptions; company name, address, and telephone number; full record with tagged fields; full record (Figure 2); company name and TRINET number; company name, SIC codes and descriptions, and all sales figures; or company name and address.

Again, contingent on the user's needs, data can be requested in a variety of formats; and data elements can be combined to assure optimum matching of information needs with available data. Due to the quarterly reloading of TRINET bases, the business professional can access the most up-to-date information on any given company.

Bibliographic data bases in DIALOG contain descriptive information in reference to a book, journal article, patent, newspaper story, dissertation, etc. Nonbibliographic data bases are of three types: 1) directory, presenting information similar to that shown in Figure 2; 2) numeric, generally including statistical tables accompanied by text; and 3) full text, i.e., data bases which include the actual and complete text of entries from encyclopedias, newspapers and dictionaries. Additional features available to searchers include the possibility of clustering groups of related data bases for preliminary searching of relevant information, and the printing of data records off-line.

In contrast to the commercial data bases is the on-line shared catalog, perhaps best exemplified by OCLC. This is a collective on-line catalog of the holdings of hundreds of libraries and other information centers around the United States and selected foreign countries. An examination of Figure 3 will show the basic data record available through a search of OCLC.

The initial section of the record includes an accession number, accompanied by formatting information that identifies through codes the type of work being described. The remainder of the data record includes further identifying information and/or descriptive cataloging information such as Library of Congress card number, International Standard Book Number (as applicable), languages, Library of Congress classification number, elements of full descriptive cataloging, notes, subject headings and added entries. In addition to the bibliographic record, the user is able to access a screen indicating which libraries own the particular title in question.

Formats included in OCLC are print formats such as books, periodicals (but not individual articles), microformats, archives and manuscripts, sheet music, machine-readable data files, etc. A variety of non-print formats are also included: audiodiscs, tapes, films, videorecordings, multimedia kits, etc. However, access to an individual data record is not as varied as it is through commercial data bases. Although subject searching is available for OCLC through Bibliographic Retrieval Service (BRS), and although OCLC is planning subject access

Figure 3

OCLC sample record

```
 Screen 1 of 2
NO HOLDINGS IN CSU - FOR HOLDINGS ENTER dh DEPRESS DISPLAY RECD SEND
 OCLC: 9645811      Rec stat: n Entrd: 830606        Used: 830625
Type: a Bib lvl: m Govt pub:   Lang:  fre Source:   Illus: a
 Repr:   Enc lvl:   Conf pub: 0 Ctry:  fr  Dat tp: s M/F/B: 00
 Indx: 0 Mod rec:   Festschr: 0 Cont:
 Desc: a Int lvl:   Dates: 1981,
  1 010     82-126245
  2 040     DLC c DLC
  3 039 0   2 b 3 c 3 d 3 e 3
  4 043     a-ch--- a e-fr---
  5 050 0   HC430.5 b .E96 1981
  6 082 0   332.6/7344/051249 2 19
  7 090      b
  8 049     CSUU
  9 245 00  Exporter, investir `a Ta"iwan : b mission d'´etude ´economique et
de prospection commerciale / c assist´ee et patronn´ee par la Soci´et´e
g´en´erale ; [publi´e sous la pr´esidence de Eric Faivre ; r´ealis´e sous la
direction de Alain Boutboul, Pascale Peres ; avec la participation de Anne
Bouchacourt et al.].

 Screen 2 of 2
 10 260 0   Paris : b Missions ´economiques, Ecole sup´erieure de commerce de
Paris, c [1981]
 11 300     161 p. : b ill. ; c 30 cm.
 12 500     Cover title.
 13 651  0  Taiwan x Economic conditions y 1975-
 14 651  0  Taiwan x Commerce z France
 15 651  0  France x Commerce z Taiwan
 16 650  0  Investments, French z Taiwan
 17 700 10  Faivre, Eric.
 18 700 10  Boutboul, Alain.
 19 700 10  Peres, Pascale.
 20 710 20  Soci´et´e g´en´erale pour favoriser le d´eveloppement du commerce
et de l'industrie en France.
 21 710 20  Ecole sup´erieure de commerce de Paris. b Missions ´economiques.
```

enhancements in the future, at the time of this writing access is primarily through author, title, Library of Congress card number, International Standard Book or Serial Number, OCLC accession number; and searches may be qualified by format type, date of publication, and region of cooperating libraries. The advantage of using OCLC searching is that a variety of formats are included in the system; and because terminals are located in many libraries, the search is fee-free to users.

A third type of available on-line system is the in-house configuration of hardware and software packages, designed for a specific purpose, containing as many fields of information as necessary or possible (contingent on the configuration), and accessed through as many indexes as the software package will allow. One example, presented in Figure 4, represents data records available through the use of dBASE III software and IBM PC hardware. The data record is from the SPIRE Project, an on-line index to musical recordings in anthologies which are part of the listening collection at Cleveland State University Libraries. The data record configuration which I designed contains fourteen fields, each of which can be altered, indexed, searched and reported. These include composer, title of cut, album title, title of collection, title of series, performing organization, conductor, solo instruments, soloists, performers, date of recording, audio disc number, label number, and date of input of the data record. The SPIRE Project record represents just one of a virtually infinite variety of records

Figure 4

SPIRE Project sample record

Record No. 49
COMPOSER

TITLE_CUT MOTETS FOR "HEC DIES": HUIC MAIN--HEC DIES; O MITISSIMA (QUANT VOI)--VIRGO--HEC DIES; DEO CONFITEMINI--DOMINO; TROP SOVENT--BRUNETE--IN SECULUM; INSTRUMENTAL MOTET: IN SECULUM

TITLEALBUM THIRTEENTH-CENTURY POLYPHONY (EARLY MEDIEVAL MUSIC)

TITLE_COLL HISTORICAL ANTHOLOGY OF MUSIC

TITLE_SER

PERF_ORGAN UN OF CHICAGO COLLEGIUM MUSICUM; UN OF KENTUCKY COLLEGIUM MUSICUM; ST. MEINRAD ARCHABBEY SCHOLA CANTORUM
CONDUCTOR BROWN, HOWARD M., MORGAN, WESLEY K.; KELLY, O.S.B., REV. COLUMBA

SOLO_INSTR

SOLOISTS

PERFORMERS

DATERECORD 00/00/75
AUDIO_NUMB AD 01699 P 249
LABEL_NUMB P249
DATE_INPUT 01/30/86

which can be created for storage and retrieval of in-house information. Although the example shown pertains to music, applications to business and other professions should be readily apparent.

Using dBASE III as a software package, one can create as many as 128 fields per data record, with the maximum record size being 4000 bytes. In addition, dBASE III allows text searching—valuable to the business executive for placing relevant notes inside a data record; and dBASE III allows for simultaneous searching of up to ten files. On-line assistance is always available, accompanied by increased speed in indexing, sorting and reporting. The system is capable of calculations, numeric precision, date conversion, statistical reporting, and a number of other features which allow efficient storage, manipulation and retrieval of information.

For the professional involved in the storage, management, access and retrieval or exchange of information on an international scale, each of the three types of data bases has its advantages. In some ways, each might serve to complement the others. Crucial, however, to the successful and optimal use of these on-line systems is recognition of the culture-imposed and language-imposed factors which govern their creation and maintenance. Subsequently, knowledge of these same cultures and/or languages is necessary for optimal retrieval of desired information. Let us look at some examples.

Central to every on-line system are the concepts of word choice (or numerical choice) and thesaurus control. Every system has built into it—or rather programmed into it—a limited word choice, and a limited thesaurus. Many of these thesauri are available in print formats; some are available on-line. To illustrate the point, consider that in the realm of business and economics, a significant body of research is published by corporate entities—sometimes institutionally affiliated—known as "schools," e.g. London School of Economics. Compare the English "school," the French "école," and the Spanish "escuela." All three words can represent the same type of corporate body indicated above. However, two problems arise in trying to search these terms in on-line systems. The first problem is that the concept of "school" is broad enough to include schools which are in fact outside the realm of business or economics. The second problem is that an on-line system does not automatically equate the English, French and Spanish equivalents, unless it has been so programmed with extensive conceptual indexing and cross-references throughout the system. In addition, consider that there are very real variations in concept of the term "school" in the English, French and Spanish cultures.

To resolve the first problem, additional terms must be added to the search in order to qualify it: terms such as "economics" or "commerce" or "business" or "finance." This was in fact what was done in the context of the search for the data record presented in Figure 1. The request made was for all data

records that combined two concepts: "école" and "commerc?", with the second term used representing a request for any word which began with "commerc" and ended in any manner—allowing for such terms as "commerce," "commercial," etc. The data base responded by informing that there were 87 records with the word "école" and 5459 records containing "commerc?"; there were, however, only 3 records which combined both terms. One of those records is represented in Figure 1.

Had I requested "school" and "commerc?", the results would have varied accordingly, altering the success of my retrieval, and possibly being more costly and less relevant. Similarly, had I requested "escuela" and "commerc?", the results would have been minimal, due to a variation in spelling; in Spanish, "commerce" is "comercio." I should therefore have requested "escuela" and "comerc?", which would allow for optimum results. Had I desired all information which dealt with English, French or Spanish schools of commerce, my options would vary. Using Boolean logic, I could request: ("school" or "école" or "escuela") and ("commerc?" or "comerc?"); or I could pursue each search independent of the other. In reality, the latter method would be more likely to yield optimum results.

In the efficient use of on-line systems, knowledge of foreign languages has many roles. It is certainly key to the accurate input of information into any system. As such, it remains important not only in the context of word choice and thesaurus control, but also in the context of varieties of indexing (e.g., string

indexing, coordinate indexing, conceptual indexing), cross-references, and any other points of access which each individual data base or system of bases might allow. If, for example, a base does not allow searching by subject, a translated title may have to be translated back to its original language, and searched accordingly.

Some bases may be capable of searching a word or term in every field which is part of the data record format; others, in contrast, require that necessary prefixes or suffixes be attached to each search term in order to optimize the search. It is important to remember that not all fields exist in all data bases; each is generated by its own producer, similar to the variations which exist in book and periodical publishing; necessarily, each will reflect editorial policy and emphases.

Other facets which come into play in on-line searching, and which reinforce the importance of knowledge of foreign languages for optimum access to information by business and other information-seeking professionals include: 1) forms and use of proper names, both individual and corporate—as well as proper place names; 2) idioms which are peculiar to a particular culture or language, especially those which are currently in vogue in any given profession, discipline or business; 3) the linguistic syntax of the culture or language in question, and its ability to be translated into an on-line syntax which can communicate with, and search, a single data base or system; 4) truncation, conjugation, variant spellings and diacritics which are part of

written or verbal language for any given culture; 5) techniques of bibliographic, textual and numerical analysis, recording and description which may vary from culture to culture and from language to language; 6) bibliographic formats and styles which are culture-bound, language-bound, and/or technology-bound; 7) on-line publications and other free text formats that are created in the context of a language system that is not the language system of the user/searcher.

Those among us who have searched on-line data bases for material written, recorded or described in a foreign language know that we are dependent on three languages simultaneously: our native language, the language in which we are seeking information, and the language of the on-line system which we are trying to access. If we are deficient in any one of these three language systems, the results of our searching will range from optimal to satisfactory to minimal; and the real problem in such deficiency and such results is that we will have failed to retrieve the same information that our colleagues and/or competitors have found with ease and strategy.

In a world where the technology of telecommunications provides us with the opportunity of tapping into data bases on a local, regional, national and international scale, the importance of knowledge of foreign languages can not be underestimated or understated. The present and future research and information needs of Academe, business, and other information-

dependent professions such as law, medicine, education, library and information science dictate that relevant information be retrieved as efficiently and as effectively as possible. Without a knowledge of foreign languages, the information-seeker stands to be shackled unnecessarily with the limitations which ignorance imposes, and with a concept of progress which will always be a little behind the times.

Grateful acknowledgment is made to the following, for permission to reproduce data records: ERIC (Educational Resources Information Center), as available through DIALOG Information Retrieval Service; TRINET COMPANY DATABASE (Trinet, Inc.), as available through DIALOG Information Retrieval Service; and OCLC (Online Computer Library Center). The SPIRE Project data record is reproduced courtesy of Cleveland State University Libraries.